



Título del trabajo: “Análisis estadístico y predicción basada en Text Mining: un caso de estudio en redes sociales”
Tutor/a: Alberto Luis Fernández Hilario
Cotutor/a: -
Departamento responsable: Departamento de Ciencias de la Computación e Inteligencia Artificial
Perfil y número de estudiantes al que va dirigido (máximo 2): 1 <i>(Rellenar sólo en caso de que la propuesta esté realizada a través de estudiante)</i>
Estudiante que propone el trabajo (Nombre, Apellidos, DNI): Sara Amaro Serrano
Tipo de trabajo: 6 <i>(consultar (*))</i>
Competencias <i>(estas son las mínimas; consultar (**) si se considera añadir otras)</i> <i>Competencias generales: G02, G03, G04, G05, G06, G07, G08, G09.</i> <i>Competencias específicas: E01, E02, E03, E04, E05, E08, E09, E10.</i>
Resultados de aprendizaje <i>(estos son los mínimos; añadir otros si se considera)</i> <ul style="list-style-type: none">• Adquirir competencias globales ligadas al desarrollo y aplicación de los conocimientos del Grado.• Adquirir competencias ligadas a la búsqueda y organización de información y documentación relevante sobre el tema objeto de estudio.• Aplicar el “pensamiento estadístico” y tener capacidad para enfrentarse a las distintas etapas de un estudio estadístico (desde el planteamiento del problema hasta la exposición de resultados).• Saber presentar, de forma escrita y oral, la memoria, los resultados y las conclusiones del trabajo realizado.
Antecedentes y resumen del tema propuesto: <p>En la actualidad, se maneja cada vez más información en formatos no estructurados o semiestructurados. Ejemplos claros son los mensajes de correo electrónico, notas de los centros de servicio al cliente, respuestas de encuestas con final abierto, fuentes de noticias, o formularios web, entre otros. Esta abundancia de información se presenta como un problema para muchas empresas a la hora de preguntarse cómo recopilar, explorar y aprovechar el conocimiento oculto detrás de esa cantidad ingente de datos.</p> <p>Entre diferentes herramientas pertenecientes al aprendizaje automático o Machine Learning, la minería de textos (en inglés “Text Mining”) es una de las soluciones más interesantes. Consiste en el proceso de analizar colecciones de materiales de texto con el objeto de capturar los temas y conceptos clave y descubrir las relaciones ocultas y las tendencias existentes sin necesidad de conocer las palabras o los términos exactos que los autores han utilizado para expresar dichos conceptos. El proceso de minería de textos incluye: identificar el texto en el que se va a realizar la minería, minar el texto y extraer datos estructurados, crear modelos de categoría y concepto y analizar los datos estructurados</p> <p>Este Trabajo Fin de Grado se va a basar y enfocar en el análisis estadístico de los datos estructurados y su posible predicción basada en Text Mining de un caso de redes sociales. Para ello, se tendrá en cuenta todo el proceso relativo a la Ciencia de Datos. El flujo de trabajo incluye desde la extracción de la información vía “Web Scrapping”, su preprocesamiento en basa a filtrado, limpieza y preparación para el aprendizaje, el entrenamiento de los modelos de extracción de conocimiento mediante herramientas estadísticas, y la representación, visualización y análisis de los resultados obtenidos.</p>



Breve descripción de las actividades presenciales y no presenciales a realizar:

Actividades presenciales (15-30%)	Planteamiento, orientación y supervisión	50 horas
	Exposición del trabajo	1 horas
	Otras:	25 horas
Actividades no presenciales (70-85%)	Preparación del trabajo	160 horas
	Elaboración de la memoria	14 horas
	Otras:	50 horas
Total (12 ECTS)		300 horas

Objetivos que se pretenden alcanzar:

- Estudio y descripción del procedimiento de Ciencia de Datos en general y Minería de Datos en particular, centrado principalmente en cada una de las etapas del trabajo.
- Extracción de los datos en crudo de redes sociales, y su posterior procesamiento.
- Estudio y aplicación de diferentes modelos estadísticos para el análisis de los datos.
- Realizar un estudio, descripción y presentación de los diferentes resultados y predicciones obtenidos durante la fase de extracción de conocimiento.
- Redacción de una memoria completa destacando los principales hitos alcanzados.

Bibliografía básica para la puesta en marcha del trabajo:

- Cathy O'Neil and Rachel Schutt. Doing Data Science . O'Reilly Media, 2013
- Scott V. Burger. Introduction to Machine Learning with R. Rigorous Mathematical Modeling. O'Reilly Media 2018 (1st Edition)
- S. García, J. Luengo, F. Herrera. Data preprocessing in Data Mining. Springer 2016 (1st Edition).
- Michael R. Berthold, Christian Borgelt, Frank Hppner, and Frank Klawonn. 2010. Guide to Intelligent Data Analysis: How to Intelligently Make Sense of Real Data (1st ed.). Springer Publishing Company
- T. Hastie, R. Tibshirani, J. Friedman. The elements of statistical Learning. Data Mining inference and prediction. 2001 Springer (2nd Edition)
- G. Shmueli N R Patel P C P Bruce Prentice Hall, Data mining for business intelligence. Wiley 2010 (2nd. edition)
- V. Cherkassky, F.M. Mulier Learning from Data: Concepts, Theory, and Methods, 2nd Edition Wiley-IEE Prees, 2007
- Atefeh, F., & Khreich, W. (2015). A survey of techniques for event detection in twitter. Computational Intelligence, 31(1), 133-164.
- Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies 5(1), -184
- Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology, 63(1), 163-173.



Tipo de trabajo (*):

1. Estudio de profundización en algún tema concreto de Estadística, o como proyecto de aplicación de la misma a estudios o problemas de otros ámbitos científicos o sociales.
2. Realización completa de todas las fases de un proyecto estadístico, bien con auxilio de prácticas en empresas o con prácticas propuestas y dirigidas por el tutor.
3. Estudio de casos, teóricos o prácticos, relacionados con la Estadística.
4. Elaboración de un informe o un proyecto de naturaleza profesional.
5. Elaboración de un plan de empresa.
6. Simulación de encargos profesionales.
7. Trabajos bibliográficos sobre el estado actual de una temática relacionada con la Estadística.
8. Creación y/o empleo de herramientas informáticas para su uso en Estadística.
9. Trabajos de inicio a la investigación.
10. Trabajos cuya finalidad sea la divulgación de la Estadística en diversos contextos.
11. Trabajos sobre Historia de la Estadística.
12. Trabajos relacionados con la docencia de la Estadística.

Competencias ()**

Competencias generales:

- G01.** Poseer los conocimientos básicos de los distintos módulos que, partiendo de la base de la educación secundaria general, y apoyándose en libros de texto avanzados, se desarrollan en la propuesta de título de Grado en Estadística que se presenta.
- G02.** Saber aplicar los conocimientos básicos de cada módulo a su trabajo o vocación de una forma profesional y poseer las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de la Estadística y ámbitos en que esta se aplica directamente.
- G03.** Saber reunir e interpretar datos relevantes para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.
- G04.** Poder transmitir información, ideas, problemas y sus soluciones, de forma escrita u oral, a un público tanto especializado como no especializado.
- G05.** Haber desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.
- G06.** Saber utilizar herramientas de búsqueda de recursos bibliográficos.
- G07.** Poder comunicarse en otra lengua de relevancia en el ámbito científico.
- G08.** Poseer habilidades y aptitudes que favorezcan el espíritu emprendedor en el ámbito de aplicación y desarrollo de su formación académica.
- G09.** Fomentar y garantizar el respeto a los Derechos Humanos, a los principios de accesibilidad universal, igualdad, y no discriminación; y los valores democráticos, de la cultura de la paz y de igualdad de género.



Competencias específicas:

E01. Conocer los fundamentos básicos del razonamiento estadístico, en el diseño de estudios, en la recogida de información, en el análisis de datos y en la extracción de conclusiones.

E02. Conocer, saber seleccionar y saber aplicar, técnicas de adquisición de datos para su tratamiento estadístico.

E03. Conocer los fundamentos teóricos y saber aplicar modelos y técnicas estadísticas en estudios y problemas reales en diversos ámbitos científicos y sociales.

E04. Saber seleccionar los modelos o técnicas estadísticas para su aplicación en estudios y problemas reales en diversos ámbitos científicos y sociales, así como conocer herramientas de validación de los mismos.

E05. Comprender la importancia de la Investigación Operativa como metodología de optimización, toma de decisiones y diseño de modelos particulares para la resolución de problemas en situaciones específicas.

E06. Comprender y utilizar básicamente el lenguaje matemático.

E07. Conocer los conceptos y herramientas matemáticas necesarias para el estudio de los aspectos teóricos y prácticos de la Probabilidad, la Estadística y la Investigación Operativa.

E08. Conocer y saber utilizar aplicaciones informáticas de análisis estadístico, cálculo numérico y simbólico, bases de datos, visualización gráfica y optimización, que sean útiles para la aplicación y desarrollo de las técnicas estadísticas.

E09. Conocer los conceptos básicos y habilidades propias de un ámbito científico o social en el que la Estadística o la Investigación operativa sean una herramienta fundamental.

E10. Tomar conciencia de la necesidad de asumir las normas de ética profesional y las relativas a la protección de datos y del secreto estadístico, como premisas que deben guiar la actividad profesional como profesionales de la Estadística.